Appendix A

Natural Gas Production Monthly Survey
By
Inderjit Kundra

The Energy Information Administration (EIA) will be initiating a new monthly survey beginning in January 2005. The purpose of this survey is to provide monthly estimates of natural gas production at the regional and national level. The regions under consideration are Texas (TX), Federal Gulf (FG), Louisiana (LA), New Mexico (NM), Oklahoma (OK), Wyoming (WY), and Others.

To accomplish this, we propose two different sample sizes corresponding to the precision levels of 5% and 1% coefficient of variations, at the national and regional levels. For this purpose, EIA-23 sampling frame for the year 2002 was used to compute the sample sizes. This frame consists of a list of 20,906 gas producing operators, of which 7285 reported a measurable production, 5,527 reported "zero" production and the remaining 8,094 reported "blank" production. For estimating the sample size, the operators showing zero or blank production were excluded from the frame. A small sample of zeros and blanks will be fielded to validate the accuracy of these responses.

Determination of sample size

To determine the sample size the Presumed Optimum Allocation formulae for sample size under optimum allocation as given by Cochran in Sampling Techniques, 2nd edition, Section 5.47, page 105 was used. To accomplish this:

- All the operators within a region were arranged in descending order of their magnitude in terms of 2002 production;
- The operators were grouped into five approximately equal size strata. The strata were created by forming a cumulative sum of production until the sum was greater than or equal to one-fifth of the total national or regional production.;
- Stratum population variances were computed, as follows:

$$S^2(h) = \frac{\sum_{i}^{N(h)} (y(hi) - (\bar{Y}(h))^2}{N(h) - 1}$$

Where:

h denotes a stratum in a region
N(h) denotes number of operators in stratum h

y(hi) denotes the production for operator i in stratum h

$\bar{Y}$ (h) denotes the mean for stratum h

- The above computed values were plugged in the following formulae to estimate the sample size.

$$CSS = \frac{[\sum_{h=1}^{5}(N(h)*S(h))]^2}{[CV*(\sum_h \sum_i y(h_i))]^2 + \sum_{h=1}^{5}(N(h)*S^2(h))}$$

Where CSS denotes calculated Sample Size
h denotes stratum in a region
N(h) denotes number of operators in a stratum
S(h) denotes standard Deviation in stratum h
CV denotes coefficient of variation
$y(h_i)$ denotes the production for operator i in stratum h

Table 1 provides the computed sample sizes corresponding to two coefficients of variation (CV) of 5% and 1% for the US, as a whole, and the 7 individual regions. This table further provides the expected number of operators to be selected with certainty (Columns 2 & 5) and the percentage of regional production covered by the certainty companies within a region (Column 3 & 6). The estimated sample sizes corresponding to respective CV of 5% and 1% are shown in columns 3 & 6.

Table1:- Estimated Sample Sizes

Region	5% CV				1% CV		
	Certainty Group	Certainty Coverage	Proposed Sample Size		Certainty Group	Certainty Coverage	Proposed Sample Size
(1)	(2)	(3)	(4)		(5)	(6)	(7)
US	**58**	**76%**	**111**		**95**	**82%**	**176**
TX	52	74%	96		134	86%	221
FG	12	72%	20		18	81%	27
LA	34	72%	57		71	87%	96
NM	15	80%	28		21	84%	36
OK	35	72%	63		56	79%	98
WY	25	93%	31		29	96%	35
OTHERS	38	73%	63		46	76%	75

Regional SUBTOTAL	**211**	**-**	**358**		**375**	**-**	**588**

It should be noted that about 41 of the certainty operators do business in more than one region. Table 2 lists these operators along with the regions they operate in. It is obvious from the last column of this table that if we count these multi-regional operators once, the overall regional sample will be reduced by 102. By counting these operators once, the anticipated regional level sample sizes corresponding to a CV of 5% and 1% will reduce to 256 and 486, respectively. This will not affect the regional sample sizes.

In addition to this survey, the monthly estimates for natural gas production for the state of Texas are developed by using data obtained from the Texas Railroad Commission (TRRC) and the Multinomial Model. This model helps in forecasting the data for the non-responding operators. Initially these estimates are treated as preliminary, and are revised as and when more data are received.

The Texas data collected from the proposed survey will be compared to the estimates based on TRRC and the multinomial model. Depending upon the results of the comparison we could recommend eliminating collection for this state.

Further, since Federal Gulf accounts for 25% of the total natural gas produced in the United States, it was decided to develop its estimates at a higher level of precision , say, a CV of 1%. This would add 7 more operators raising its sample size from 20 to 27. Using this option, the overall sample size will increase to 263.

Two sample selection plans will be used to select this sample. The first plan will select this sample using a single stage stratified random sampling plan with operator as the ultimate sampling unit. This plan selects the sampling units within a stratum with equal probability.

The second plan will select the sample using a probability proportional to size selection plan. Under this plan sampling units will be selected systematically, with probability proportional to their measure of size.

The estimates along with their variances will be compared to evaluate the efficiency of the two plans. This will help EIA to decide which sampling plan to adopt. The results of these manipulations will be compared at least for two previous years.

EIA has selected a sample of 200 of the operators that are recorded as reporting “zero” in the frame, and also has selected a sample of 200 of the operators that are recorded as reporting “blank”. EIA’s Dallas Field Office will evaluate the operators in the sample and provide all available information. If questions remain about their activity, or possible activity these operators will be contacted.

The Committee is requested to comment on all aspects of the proposed methodology and provide guidance for conducting this survey. The Committee is also invited to comment on how to take care of the operators with zero or blanks, especially, if our samples of the zeros and blanks result into a few operators having large production.

Table 2:- Multi-State Reporting Adjustment

FG **(12/19)**	LA **(23/35)**	NM **(15/16)**	OK **(22/36)**	TX **(32/53)**	WY **(17/26)**	OTHER **(22/39)**	Operator Name	# of Regions	Excessive Counts
FG	LA	NM	OK	TX	WY	OT	CHEVRON U S A	7	6
FG	LA	NM	OK	TX		OT	EL PASO	6	5
FG	LA		OK	TX	WY	OT	ANADARKO	6	5
	LA	NM	OK	TX	WY	OT	BURLINGTON	6	5
	LA	NM	OK	TX	WY	OT	X T O ENERGY INC	6	5
FG	LA	NM	OK	TX			BP PLC	5	4
FG	LA		OK	TX		OT	EXXON MOBIL	5	4
FG	LA	NM		TX	WY		DEVON ENERGY	5	4
FG	LA	NM	OK	TX			APACHE CORP	5	4
		NM	OK	TX	WY	OT	MARATHON OIL CO	5	4
		NM	OK	TX	WY	OT	E O G RESOURCES	5	4
	LA	NM	OK	TX		OT	CONOCOPHILLIPS	5	4
FG				TX	WY	OT	SHELL OIL CO	4	3
FG	LA		OK	TX			NEWFIELD	4	3
FG	LA		OK	TX			KERR MCGEE O&G C	4	3
	LA			TX	WY	OT	MOBIL	4	3
		NM	OK	TX		OT	OCCIDENTAL	4	3
	LA		OK	TX	WY		SAMSON	4	3
	LA		OK		WY	OT	CABOT OIL & GAS	4	3
FG	LA					OT	UNOCAL CORP	3	2
FG	LA					OT	AMERADA HESS	3	2
				TX	WY	OT	WESTPORT	3	2
			OK	TX		OT	DOMINION	3	2
	LA		OK	TX			OCEAN ENERGY INC	3	2
		NM			WY	OT	WILLIAMS	3	2
	LA			TX			NOBLE ENERGY INC	2	1
	LA			TX			HUNT OIL CO	2	1
		NM		TX			PURE RESOURCES	2	1
				TX		OT	CALPINE NATURAL	2	1
				TX		OT	PIONEER NATURAL	2	1
			OK	TX			CHESAPEAKE	2	1
	LA			TX			HILCORP ENERGY	2	1
	LA			TX			HUNT PETROLEUM	2	1
			OK	TX			CIMAREX ENERGY	2	1
			OK	TX			GRUY PETROLEUM	2	1
	LA			TX			K C S RESOURCES	2	1
		NM				OT	ENERGEN	2	1
			OK		WY		QUESTAR	2	1
					WY	OT	BROWN TOM INC	2	1
		NM			WY		YATES PETROLEUM	2	1

Total Number of Excessive Counts	102

.